Creation Method of Fuzzy Modeling with Variation Degree

Michiharu Maeda[1] Lixin Ma[2] Hiromi Miyajima[3]

1 Kurume National College of Technology
Kurume, 830-8555 Japan

2 Shanghai University of Electric Power
Shanghai, 200090 China
-
3 Faculty of Engineering, Kagoshima University
Kagoshima, 890-0065 Japan

Abstract: - This paper presents a novel method of rule creation in fuzzy modeling with variation degree. The present method has a construction mechanism of the rule unit that is applicable in a parameter for the central value of the membership function in the antecedent part. The approach is to create the rule unit near the position of central value whose unit has the largest degree of variations in fuzzy reasoning. As rules are generally needed at the location of the large amount of variation, fuzzy reasoning appropriately advances. Experimental results are presented in order to show that the present method is effective on the inference error and the number of learning iterations.

Key-Words: - Fuzzy modeling, Creation method, Inference error, Gradient descent, Learning Variation degree, Numerical experiment

1 Introduction

With the object of automatically constructing fuzzy inference rules utilizing the learning process, a number of approaches have been studied on fuzzy modeling [1], [2]. They are aimed at drastically reducing the processing labor by applying the learning function to the tuning of fuzzy inference rules [3]–[9]. In fuzzy modeling, when the rule is constructed on the fuzzy system, it is important to minimize the inference error, to lighten the rule number, and to shorten the learning process. With respect to the constitution of fuzzy inference rules, the reduction methods [10] have been introduced in order to minimize the inference error and to shorten the learning process. In the techniques, the rule which seems to be little influence on the inference error is removed, and the methods proceed with learning by gradient descent. Thus, the extraction of appropriate rules become possible by reducing the number of redundant rules. However, when the reduction techniques were adopted, the different results were shown according to the application of various functions.

In this paper, we present a novel method of rule construction in fuzzy modeling by using gradient descent. The present approach is described with a construction method applicable to a parameter for the central value of the membership function in the antecedent part. The technique is to create the rule unit near the position of central value whose unit has the largest degree of variations in fuzzy reasoning. As rules are generally needed at the location of the large amount of variation, fuzzy reasoning appropriately advances. Experimental results are presented in order to show that the validity of the present method is confirmed.

2 Fuzzy Reasoning and Self-Tuning Methods

The simplified fuzzy reasoning which treats the consequent part as the reasoning method at the real number is used. When the input is (x_1, x_2, $\cdots$, x_n) and the output is y, the procedure is expressed as follows:

$$R^i: \quad \text{If } x_1 \text{ is } M_{i1} \text{ and } \cdots \text{ and } x_n \text{ is } M_{in} \text{ then } y \text{ is } w_i,$$

where M_{ij} is a membership function in the antecedent part, and w_i is a real number in the consequent part. The membership function M_{ij} in the antecedent part is set separately for every rule, and has the index i for the rule number. The membership function M_{ij} in the antecedent part is an isosceles triangle, and can be expressed using the central value a_{ij} and the width b_{ij} in the following equation.

$$M_{ij}(x_j) = \begin{cases} 1 - \frac{2|x_j - a_{ij}|}{b_{ij}} & (|x_j - a_{ij}| \le b_{ij}/2) \\ 0 & \text{(otherwise).} \end{cases} \quad (1)$$

The membership value μ_i of the i-th rule is obtained by the following equation.

$$\mu_i = \prod_{j=1}^{n} M_{ij}(x_j). \quad (2)$$

Therefore, the reasoning result y consists of

$$y = \frac{\sum_{i=1}^{\gamma} \mu_i w_i}{\sum_{i=1}^{\gamma} \mu_i}. \quad (3)$$

The function which shows the shape of the membership function is adjusted by the delta rule with the central value a_{ij}, the width b_{ij}, and the actual value w_i in the consequent part. It is possible to consider the delta rule as a minimization problem of the objective function E which shows the error between the output value (i.e., y of fuzzy reasoning) of fuzzy systems and the desired output value y_r, as shown in the following equation.

$$E = \frac{1}{2}(y - y_r)^2. \quad (4)$$

In order to decrease the value of the objective function E, when input-output data $(x_1, \cdots, x_n, y_r)$ of the $n+1$ dimension are given, the gradients ($\partial E/\partial a_{ij}$, $\partial E/\partial b_{ij}$, $\partial E/\partial w_i$) of objective function E are calculated on a_{ij}, b_{ij}, and w_i. Subsequently, the values of a_{ij}, b_{ij}, and w_i are updated according to

$$\Delta a_{ij} = -\eta_a \frac{\partial E}{\partial a_{ij}} \quad (5)$$

$$\Delta b_{ij} = -\eta_b \frac{\partial E}{\partial b_{ij}} \quad (6)$$

$$\Delta w_i = -\eta_w \frac{\partial E}{\partial w_i}, \quad (7)$$

where η_a, η_b, and η_w are learning constants. $\partial E/\partial a_{ij}$, $\partial E/\partial b_{ij}$, and $\partial E/\partial w_i$ are calculated as

$$\frac{\partial E}{\partial a_{ij}} = \frac{\mu_i}{\sum_{i=1}^{\gamma} \mu_i}(y - y_r)(w_i - y) \cdot \text{sgn}(x_j - a_{ij}) \frac{2}{b_{ij} M_{ij}(x_j)} \quad (8)$$

$$\frac{\partial E}{\partial b_{ij}} = \frac{\mu_i}{\sum_{i=1}^{\gamma} \mu_i}(y - y_r)(w_i - y) \cdot \frac{1 - M_{ij}(x_j)}{b_{ij} M_{ij}(x_j)} \quad (9)$$

$$\frac{\partial E}{\partial w_i} = \frac{\mu_i}{\sum_{i=1}^{\gamma} \mu_i}(y - y_r), \quad (10)$$

where

$$\text{sgn}(\theta) = \begin{cases} -1 & (\theta < 0) \\ 0 & (\theta = 0) \\ 1 & (\theta > 0). \end{cases}$$

By giving input-output data one after another and repeating the learning process, the shape of the membership function in which the value of objective function E becomes minimal is determined. The adjustment of the shape of the membership function is carried out until the inference error $D(t)$ shown in the following equation is less than the desired value δ, for the given input-output data (x_1^p, x_2^p, $\cdots$, x_n^p, y_r^p), $p = 1, 2, \cdots, P$.

$$D(t) = \frac{1}{P}\sum_{p=1}^{P}(y^p - y_r^p)^2, \qquad (11)$$

where y^p is an output of fuzzy reasoning.

3 Creation Method and Algorithm

In this section, the creation method and algorithm are described. A new rule of the present method is constructed near the position of central value whose unit has the largest degree of variations in fuzzy reasoning. Generally, rules are needed at the location of the large amount of variation. Therefore we define the positions χ_{ij-} and χ_{ij+} from the current central value a_{ij} as follows:

$$\begin{aligned} \chi_{ij-} &= a_{ij} - \varepsilon \\ \chi_{ij+} &= a_{ij} + \varepsilon, \end{aligned} \qquad (12)$$

where ε is a minute constant.

Furthermore, we define $f(x)$ according to Eqs. (1), (2), and (3) as follows:

$$f(x) \equiv y(x_j = x). \qquad (13)$$

Therefore, the function $f(x)$ is the reasoning result for the given input x. Using χ_{ij-} and χ_{ij+}, the amount of variations $f'(a_{ij})$ for the central value a_{ij} is calculated as

$$f'(a_{ij}) = \frac{f(\chi_{ij+}) - f(\chi_{ij-})}{|\chi_{ij+} - \chi_{ij-}|}. \qquad (14)$$

In order to obtain the absolute value of $f'(a_{ij})$, the following equation is introduced.

$$\mathcal{F}(a_{ij}) = \cos f'(a_{ij}). \qquad (15)$$

$\mathcal{F}(a_{ij})$ represents the degree of variations for the inference value in the central value a_{ij}, and becomes the criterion of rule creation. Thus, a new rule is created after the central value a_{ij} is selected as a candidate of rule creation when $\mathcal{F}(a_{ij})$ has the minimum (i.e., the reasoning result is the most changeful).

Furthermore, so as not to gather around local positions for central values, a rule is created when the distribution of central values is thin. The judgment of whether it is thin or not will be carried out according to a value $\mathcal{D}(a_{ij})$ which includes both of the distance between the noticed central value and its first-nearest value, and the distance between the noticed value and its second-nearest value. Thus, the following function is required:

$$\frac{2}{\gamma - 1} < \mathcal{D}(a_{ij}), \qquad (16)$$

where γ is a rule number.

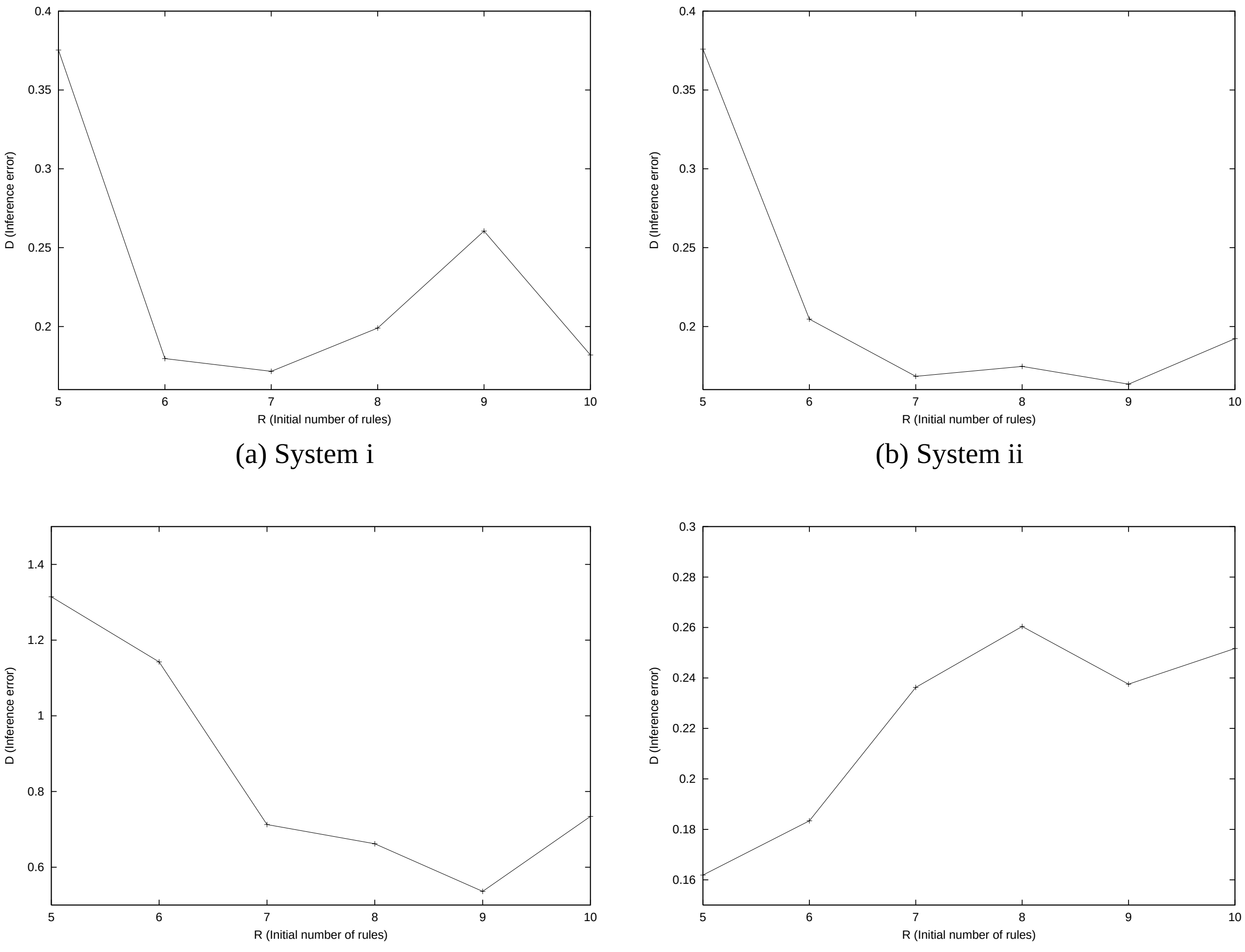
Under the condition of this function, a new rule is created as the position $a_{ij} + \varepsilon$ near the selected rule satisfied in Eq. (15). The central value of the chosen rule become the position $a_{ij} - \varepsilon$.

By adapting the creation standard, the rules are created in which the position of the most changeful in the reasoning. As the rule is constructed for the position with the necessity, fuzzy reasoning appropriately advances, given in the next section.

[*Creation algorithm*]

Step A1 Initialization:
Give central value a_{ij} and width b_{ij} in the antecedent part, real number w_i in the consequent part, creation threshold δ_c, termination threshold δ_T, maximum number of learning iterations T_{max}, initial rule number R_I, and final rule number R_T. Set $t \leftarrow 0$ and $\gamma \leftarrow R_I$.

Step A2 Self-tuning:
(A2.1) Let $p = 1$.
(A2.2) Let s_p be an index selected at random among $\{1, 2, \cdots, P\}$, for all $s_i \neq s_j$.
(A2.3) Allot the input-output data ($x_1^{s_p}$, $x_2^{s_p}$, $\cdots$, $x_n^{s_p}$, $y_r^{s_p}$).
(A2.4) Derive the output of fuzzy inference y^p performed by the simplified fuzzy reasoning.
(A2.5) Adapt w_i according to Eq. (7) and repeat the fuzzy reasoning at A2.4.



(a) System i (b) System ii

(c) System iii (d) System iv

Figure 1: Relation between inference error ($\times 10^{-4}$) and initial number of rules for the creation model. The results are averages of 1000 trials. When $R = 10$, the creation model is equivalent to the conventional model because there are no existing rules to create.

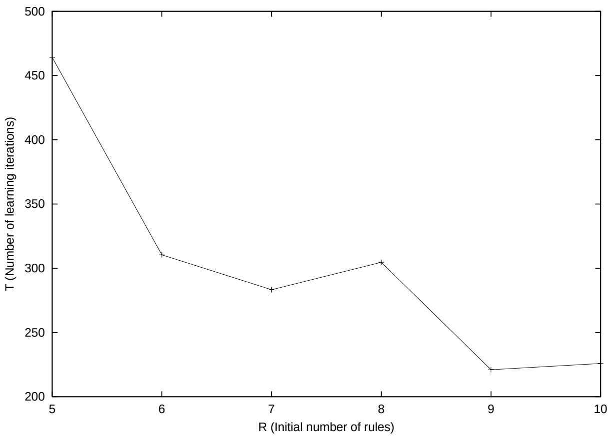
(A2.6) Adapt a_{ij} and b_{ij} of the membership functions in the antecedent part, according to Eqs. (5) and (6), respectively.
(A2.7) If $p < P$, then set $p \leftarrow p+1$ and go to A2.2, otherwise set $t \leftarrow t+1$ and go to Step A3.

Step A3 Rule creation:
(A3.1) Calculate $D(t)$ and $\Delta D(t)$ according to Eqs. (11) and (12), respectively.
(A3.2) If $\gamma < R_T$ and $\Delta D(t) \leq \delta_c$, then go to A3.4.
(A3.3) If $\gamma = R_T$, then go to Step A4, otherwise go to Step A2.
(A3.4) Create the k-th rule according to the creation method. Set $\Re \leftarrow \Re + \{k\}$ and $\gamma \leftarrow \gamma + 1$. Go to Step A2.

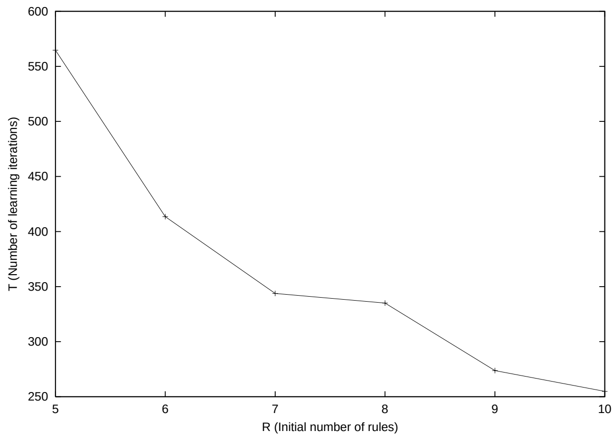
Step A4 Termination condition:
If $t = T_{max}$ or $\Delta D(t) \leq \delta_T$, then terminate, otherwise go to Step A2.

At steps A3.1 and A4, the value $\Delta D(t)$ is calculated as follows.

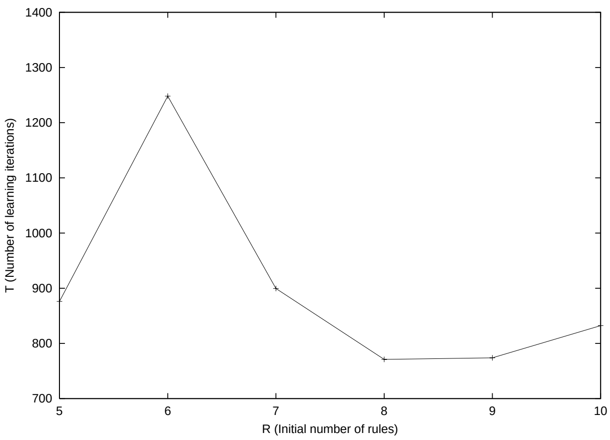
$$\Delta D(t) = \frac{|D(t) - D(t-1)|}{D(t-1)} \quad (17)$$



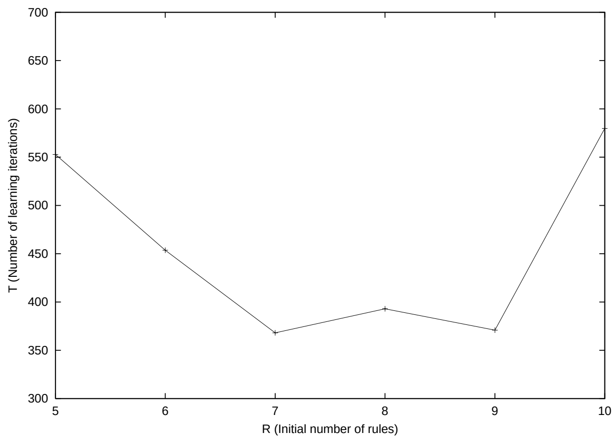
(a) System i



(b) System ii



(c) System iii



(d) System iv

Figure 2: Relation between number of learning iterations and initial number of rules for the creation model. The results are averages of 1000 trials. When $R = 10$, the creation model is equivalent to the conventional model because there are no existing rules to create.

4 Numerical Experiments

We perform the function approximation by using the above model. The systems are identified by the data as fuzzy inference rules, with the utilization of input-output data from the known function as follows:

$$\text{(i)} \quad y = \frac{\cos \pi x + 1}{2}$$

$$\text{(ii)} \quad y = \frac{\sin \pi x + 1}{2}$$

$$\text{(iii)} \quad y = \begin{cases} -\sin \pi x & (x \leq 0) \\ \sin \pi x & (0 < x) \end{cases}$$

$$\text{(iv)} \quad y = \begin{cases} -x^3 & (x \leq 0) \\ x & (0 < x) \end{cases}$$

The domain of each variable x and output y_r normalize within $[-1, 1]$ and $[0, 1]$, respectively. The parameters are chosen as follows: $\eta_a = 0.1$, $\eta_b = 0.1$, $\eta_w = 0.2$, $\varepsilon = 0.01$, $P = 100$, $\delta_c = 10^{-2}$, $\delta_T = 10^{-4}$, $T_{max} = 100000$, and $R_T = 10$.

Figure 1 shows the influence of the inference error on the initial number of rules for each model. For system i, as the initial number of rule is 6 and 7, the inference error is smaller than the others. For system ii, the inference error is small compared to the conventional model ($R = 10$)

when the initial number of rules is 7, 8, and 9. Especially, the inference error is the best when the initial number of rule is 9. For system iii, as well as system ii, the inference error is the better than that of the conventional model when the initial number of rule is 7, 8, and 9. The inference error is the best when the initial number of rule is 9. System iv exhibits good results when the initial number of rule is 5, 6, 7, and 9. In this system, it is proven that the best result shows when the initial number of rule is 5. The effectiveness differs among the present systems according to the inference error.

Figure 2 shows the relation between number of learning iterations and initial number of rules for the creation model. There seems to be a general trend that the number of learning iterations increases with a greater initial number of rules. However system iii and iv exhibit various trends. For systems iii and iv, the number of learning iterations gives small value when the initial number of rule is 8 and 7, respectively. The effectiveness also differs among the present systems according to the number of learning iterations.

5 Conclusions

In this paper, we have presented a novel construction method of fuzzy inference rules with variation degree and have examined its validity through numerical experiments. The approach was the creation mechanism of rule unit that was applicable in a parameter for the central value of the membership function in the antecedent part. The result of numerical experiments was that the present model led to different effects according to the application of various functions. For the future works, we will study more effective approaches.

References

[1] R.R. Yager and L.A. Zadeh, Fuzzy sets, neural networks, and soft computing, *Van Nostrand Reinhold*, 1994.

[2] N.K. Kasabov, Foundations of neural networks, fuzzy systems, and knowledge engineering, *The MIT Press*, 1996.

[3] H. Ichihashi and T. Watanabe, Learning control by fuzzy models using a simplified fuzzy reasoning, *Journal of Japan Society for Fuzzy Theory and Systems*, vol.2, 1990, pp.429–437.

[4] H. Takagi, T. Kouda, and Y. Kojima, Neural-networks designed on approximate reasoning architecture, *Journal of Japan Society for Fuzzy Theory and Systems*, vol.3, 1991, pp.133–141.

[5] N. Nomura, I. Hayashi, and N. Wakami, A self-turning method of fuzzy reasoning by delta rule and Its application to a moving obstacle avoidance, *Journal of Japan Society for Fuzzy Theory and Systems*, vol.4, 1992, pp.379–388.

[6] H. Miyajima, S. Fukumoto, and Y. Ishizuka, Fuzzy modeling based on deleting rules, *IEICE Trans.*, vol.J77-A, 1994, pp.1555–1562.

[7] H. Miyajima, K. Kishida, and S. Fukumoto, Constructive, destructive and simplified learning methods of fuzzy inference, *IEICE Trans. Fundamentals*, vol.E78-A, 1995, pp.1331–1338.

[8] K. Kishida, H. Miyajima, and M. Maeda, Destructive fuzzy modeling using neural gas network, *IEICE Trans. Fundamentals*, vol.E80-A, 1997, pp.1578–1584.

[9] S. Fukumoto, H. Miyajima, K. Kishida, and Y. Nagasawa, An investigation of fuzzy model using AIC, *IEICE Trans. Fundamentals*, vol.E80-A, 1997, pp.1698–1703.

[10] M. Maeda and H. Miyajima, Fuzzy modeling in some reduction methods of inference rules, *IEICE Trans. Fundamentals*, vol.E84-A, 2001, pp.820–828.